Stats 2MB3, Tutorial 1

Jan 16, 2015

TA Information

- Name: Chengwei Qin (PhD student in Statistics)
- E-mail: qinc4@math.mcmaster.ca
- Webpage: http://ms.mcmaster.ca/~qinc4/
- Office Hours: 10am to 12pm, Friday, HH401

Data Description/Visualization

What u gonna do if one day you receive this file

3	2013/9/3	1057488	2	3	NA	1	2013/9/3	2013/9/4	2013/9/4	NA	NA	NA
4	2013/9/3	1057490	9	0	2013/9/3	1	2013/9/3	2013/9/4	2013/9/4	NA	########	2013
5	2013/9/3	1057492	2	4	NA	1	2013/9/3	2013/9/3	2013/9/3	2013/9/3	NA	NA
6	2013/9/3	1057496	108	630	2013/9/3	20	2013/9/3	2013/9/3	2013/9/5	2013/9/5	2013/9/6	2013
7	2013/9/3	1057498	2	8	2013/9/3	2	2013/9/3	2013/9/3	2013/9/3	2013/9/3	2013/9/3	2013
8	2013/9/3	1057502	446	1736	NA	27	2013/9/3	2013/9/3	2013/9/4	2013/9/3	2013/9/4	NA
9	2013/9/3	1057506	262	531	########	24	2013/9/3	2013/9/3	NA	2013/9/3	########	2013
10	2013/9/3	1057514	128	651	2013/9/3	31	2013/9/3	2013/9/3	2013/9/3	2013/9/3	2013/9/5	2013
11	2013/9/3	1057516	23	109	2013/9/3	8	2013/9/3	2013/9/3	2013/9/3	2013/9/3	NA	NA
12	2013/9/3	1057518	2	0	NA	1	2013/9/3	2013/9/4	2013/9/4	NA	NA	NA
13	2013/9/3	1057520	1	1	NA	1	2013/9/3	2013/9/3	2013/9/3	NA	NA	NA
14	2013/9/3	1057522	1	0	NA	1	NA	NA	NA	NA	NA	NA
15	2013/9/3	1057524	2	17	NA	3	2013/9/3	2013/9/3	2013/9/3	2013/9/3	NA	2013
16	2013/9/3	1057526	1	4	NA	1	2013/9/3	2013/9/3	2013/9/3	NA	2013/9/3	NA

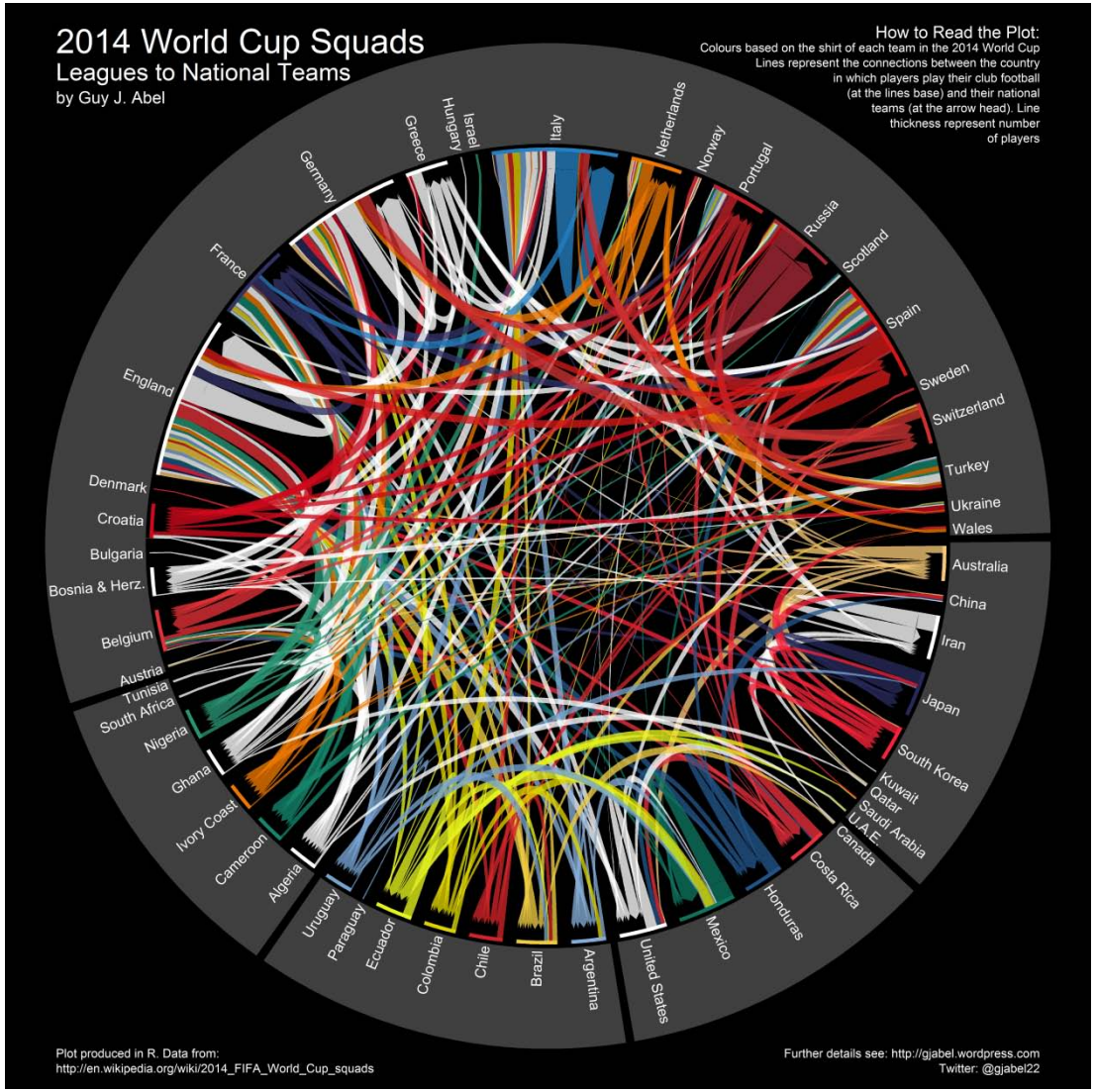
⋮

299988	2013/12/12	1874052	1	0	NA	1	NA	NA	NA	NA	NA	NA
299989	2013/12/12	1874054	7	2	NA	1	########	########	########	########	########	NA
299990	2013/12/12	1874058	264	884	########	49	########	########	########	########	########	##
299991	2013/12/12	1874066	1	0	NA	1	NA	NA	NA	NA	NA	NA
299992	2013/12/12	1874070	1	0	NA	1	NA	NA	NA	NA	NA	NA
299993	2013/12/12	1874074	3	0	NA	1	NA	NA	NA	NA	NA	NA
299994	2013/12/12	1874076	49	101	NA	14	########	########	NA	########	########	##
299995	2013/12/12	1874078	112	252	NA	18	########	########	########	########	########	NA
299996	2013/12/12	1874082	1	0	NA	1	NA	NA	NA	NA	NA	NA
299997	2013/12/12	1874084	2	5	NA	1	########	########	########	########	########	##
299998	2013/12/12	1874086	4	3	NA	1	########	########	NA	NA	NA	NA
299999	2013/12/12	1874090	2	8	NA	1	########	########	########	########	########	NA
300000	2013/12/12	1874094	1	0	NA	1	NA	NA	NA	NA	NA	NA
300001	2013/12/12	1874102	6	26	NA	5	########	########	########	########	########	NA

Global Connections on Facebook

1.3 billion active Facebook users by July 1st, 2014

2014 World Cup Squad



Stem-and-leaf Displays

1. Take leading digits as stem values and trailing digits as leaves;
2. Vertically list stem values;
3. Record leaves beside corresponding stems;
4. Indicate the units.

Ex 10, page 24

Consider the strength data for beams given in Example 1.2

a. Construct a stem-and-leaf display of the data. What is representative strength value? Observations highly concentrate on it or not?

b. Is the display symmetric about the representative value or how would you describe it?

c. Any outlying strength values?

d. What proportion of observations exceed 10 Mpa?

```
> data<-c(5.9,7.2,7.3,6.3,8.1,6.8,7.0,7.6,6.8,6.5,7.0,6.3,7.9,9.0,8.2,8.7,7.8,9.7,7.4,7.7,9.7,7.8,7.7,11.6,11.3,11.8,10.7)
> stem(data)

  The decimal point is at the |

   5 | 9
   6 | 33588
   7 | 00234677889
   8 | 127
   9 | 077
  10 | 7
  11 | 368
```

a. A representative strength value would be around 7.5. For the most part, the observations concentrate on stem 7.

b. The display appears to be skewed to smaller values.

c. No outlier.

d. 10.7, 11.3, 11.6, 11.8 are larger than 10. The proportion is 4/27.

Ex 28, page 27

- Human measurements provide a rich area of application for statistical methods. The article reported on a study of children talking to themselves. It was thought that private speech would be related to IQ, because IQ is supposed to measure mental maturity, and it was known that private speech decreases as students progress through the primary grades. The study included 33 students whose first-grade IQ scores are given here:

Ex 28, page 27

```
> data<-c(82,96,99,102,103,103,106,107,108,108,108,108,109,110,110,111,113,113,113,113,115,115,118,118,119,121,122,122,127,132,136,140,146)
> stem(data)

  The decimal point is 1 digit(s) to the right of the |

   8 | 2
   9 | 69
  10 | 2336788889
  11 | 001333355889
  12 | 1227
  13 | 26
  14 | 06
```

The data is fairly symmetrical, and the representative value is about 113.

Measures of Variability

Sample variance:

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}$$

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

Ex 46, page 43

- The accompanying observations on stabilized viscosity for specimens of a certain grade of asphalt with 18% rubber added are from the article " Viscosity Characteristics of Rubber-Modified Asphalts": 2781, 2900, 3013, 2856, 2888.
- a) What are the values of the sample mean and sample median?
- b) Calculate the sample variance using the computational formula.

Ex 46, page 43

a)

mean=(2781+2900+3013+2856+2888)/5=2887.6

Since 2781<2856<2888<2900<3013, the median is 2888.

b)

After subtracting 2700, we have {81, 200, 313, 156, 188}. By using the sample variance equation (degree of freedom n-1), the variance is equal to 7060.3.